基于随机森林算法的土壤含盐量预测

李小雨， 贾科利， 魏慧敏， 陈睿华， 王怡婧

（宁夏大学地理科学与规划学院，宁夏 银川 750021）

摘 要： 快速监测区域土壤盐渍化信息，对于盐渍化治理与生态环境保护具有重要意义。本文以Sentinel-2A和Landsat8 OLI遥感影像为数据源，以银川平原为研究区，利用谷歌地球引擎（Google Earth Engine，GEE）平台，基于随机森林算法，通过建立光谱指数特征与地面实测土壤含盐量之间的关系，进行土壤含盐量估算。结果表明：GEE能够为土壤含盐量预测提供可靠的数据支撑；以Sentinel-2A为数据源建立的随机森林模型具有更好的预测精度（R^2=0.789，RMSE=1.487），优于Landsat8 OLI，可用于土壤含盐量高分辨率遥感估算，能够为大尺度土壤含盐量监测工作提供理论支撑。

关键词： 土壤含盐量；Google Earth Engine；随机森林；预测；银川平原

土壤盐渍化是全球干旱半干旱地区面临的重要的生态环境问题[1]。全世界受盐碱化危害的面积近10×10^8 hm^2，中国盐渍土面积已超过3.3×10^7 hm^2，并以每年约1.5×10^6 hm^{2}的速度增加[2-4]。受到全球气候变暖及人类活动的影响，土壤盐渍化已成为限制农业发展和资源可持续利用的重要因素，是导致土地退化的主要因素，严重影响了区域生态环境稳定、粮食安全和农业的可持续发展[5]。因此，快速、精准、大面积监测土壤盐渍化对农业生产及生态环境保护意义重大[6]。

传统监测土壤盐分的方法耗时、费力，遥感技术作为一种新的监测方法，相比于传统方法在进行大尺度监测研究方面具有很大的优势[7-9]，在全球和区域土壤盐渍化监测中发挥着重要作用[10-11]。随着遥感技术的发展，不同来源遥感数据为土壤盐渍化监测提供了数据支持，而建立土壤盐渍化信息提取的方法，是快速获取盐渍化信息的关键。近年来，随机森林算法以其优越性、高效性、稳健性和精确性正逐步应用于土壤盐分估算中，通过建立实测土壤盐分与土壤的光谱信息间复杂的非线性关系，充分利用传感器获得的光谱信息，提高了估算精度[12-14]。厉彦玲等[15]将环境一号卫星HSI高光谱影像与Landsat8 OLI多光谱影像进行融合，利用机器学习模型反演黄河三角洲地区土壤盐分，通过对比发现，随机森林模型取得较高的精度；张智韬等[16]借助无人机多光谱遥感技术监测内蒙古河套灌区沙壕渠灌域的土壤盐分，证明随机森林算法优于支持向量机算法和BP神经网络算法，且能够在短时间内进行大范围盐分监测；Fathizad等[17]使用有限的实地测量数据基于随机森林算法预测伊朗中部沙漠的土壤盐分时空分布，获得了最佳的预测结果。随机森林模型的使用，提升了人们对于土壤盐渍化变化及预测的精度，丰富了盐渍化研究在时空方面的意义。

在平台应用上，谷歌地球引擎（Google Earth Engine，GEE）平台集数据获取、存储、处理和分析于一体，可以高效地实现海量影像数据的预处理及计算分析[18]。国内外已经有专家学者借助该平台开展土地利用及农作物分类算法[19-20]、作物产量估算[21-22]、生物量估算[23]、水体及森林变化动态监测[24-25]、城市蔓延[26]等研究。上述研究表明，GEE云平台可以高效地实现海量影像数据的预处理及计算分析，但已有研究很少尝试用GEE云平台进行土壤盐分的估算。

收稿日期： 2023-01-06； **修订日期：** 2023-06-12
基金项目： 国家自然科学基金项目（42061047，42067003）；宁夏回族自治区重点研发计划项目（2021BEG03002）；国家重点研发计划项目（2021YFD1900602）
作者简介： 李小雨（1997-），女，硕士研究生，主要从事遥感监测与分析研究. E-mail: lixiaoyu032@163.com
通讯作者： 贾科利. E-mail: jiakl@nxu.edu.cn

基于随机森林算法估算精度高、GEE云平台高效的优势，在GEE云计算平台，使用该平台自带的JavaScript API在线编辑器(https://code.earthengine.google.com/)编程，以宁夏银川平原为例，利用Sentinel-2A和Landsat8 OLI卫星遥感影像，基于光谱指数建立随机森林模型，估算土壤含盐量(Soil salt content，SSC)，以期丰富随机森林算法和GEE云平台在土壤盐渍化监测中的应用，为灌区土壤盐渍化信息的准确预测和监测提供技术支持。

1 研究区概况与数据来源

1.1 研究区概况

银川平原(37.83°~39.38°N，104.28°~107.65°E)地处宁夏回族自治区北部(图1)，南北长约165 km，东西宽10~50 km，面积约7615 km²。该区地处黄河灌溉中上游，处于贺兰山东麓的洪积扇与平原之间，是黄河流域生态保护和高质量发展的先行区，更是西北干旱与半干旱地区土壤盐渍化的典型样区。研究区属于温带干旱区，具有暖温带季风气候，年平均气温9 ℃，年均降水量在150~203 mm，年日照时数2500~3100 h，年均蒸发量>1825 mm。银川平原主要是由冲洪积平原构成，地势低洼、排水条件差、地下水埋深浅、蒸发强、水盐分集中、地势高、灌溉不合理，是宁夏土壤盐渍化最严重的地区之一。主要作物有水稻、小麦、玉米。

1.2 土壤盐分数据获取

本文以5 km×5 km的格网设置采样点，采样时每个样点采用梅花形采样法，在采样点30 m×30 m范围内，采集0~20 cm土壤混匀，运用四分法留约500 g土壤样品装进密封袋带回，同时记录采样点的编号、坐标信息、土地利用、植被长势等信息。土壤样品采集于2021年3月和2022年4月，2021年采集166个样点，2022年采集133个样点，共采集299个样点(图1)。将所有土壤样本剔除荒草和碎石块等杂质后自然风干，研磨后过2 mm筛，以水土比5:1配置提取液，用电导率法[27-28]测算含盐量，删除异常样点后得到0~20 cm表层土壤样本共297份。

1.3 数据处理

为消除不同指标之间维度的影响，通过GEE云平台调用normalization函数，将数据进行归一化处理，将所有输入数据归一化到0~1，归一化公式为：

$$x_* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

式中：x_i为该点实测样本值；$x_{\min}$为最小样本；$x_{\max}$为最大样本；max和min是给定缩放范围的最大值和最小值；x_*为归一化后的数据。

1.4 遥感影像获取与处理

研究使用的Landsat8 OLI Level-2和Sentinel-2 Level-2A地表反射率数据来源于GEE云平台(https://code.earthengine.google.com/)。数据基准面为WGS84，地图投影为UTM投影，Landsat8 OLI的空间

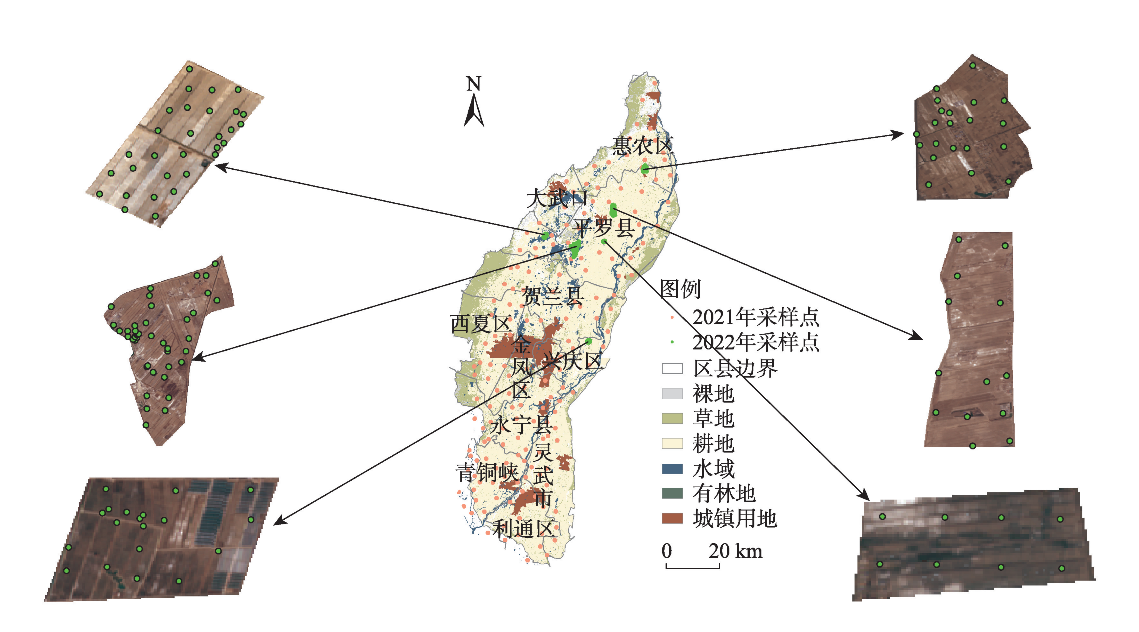


图1 研究区位置和采样点分布

Fig. 1 Location and distribution of sampling points of study area

分辨率为30 m，Sentinel-2A的空间分辨率为10 m，调用filter函数筛选云量小于10%的影像。数据经过几何校正、辐射校正和大气校正。由于数据中均集成了用于影像质量评估的QA波段，利用QA波段提供的云掩膜信息可进行遥感影像的去云操作。

影像成像时间范围分别设置为2021年3月1—31日和2022年4月1—30日，与野外采样时间相对应。由于在GEE云平台进行影像镶嵌的过程中容易出现影像重叠和不连续的问题，故根据设定的时间范围，调用median函数计算影像集的中位数，筛选与该中位数影像差异最小的影像进行镶嵌[29]。最后，选择影像中的Blue、Green、Red、NIR、SWIR1和SWIR2六个常用的波段进行影像波段合成，在合成影像上利用研究区矢量边界裁剪得到研究区影像数据，以上操作均基于GEE云平台通过JavaScript语言实现。

2 研究方法

2.1 光谱指数选取

在干旱半干旱地区，光谱指数是监测土壤盐渍化的有效方法[30]。已有研究表明，学者们在选取光谱指数时，常用的是植被指数和盐分指数。由于植被指数难以准确反映稀疏植被地区的光谱信息，基于此产生了诸多基于植被指数与盐分指数组合反演土壤盐渍化的研究[31-32]。考虑到研究区处于干旱半干旱区，本文综合了植被指数和盐分指数两种光谱指数。盐分指数包括：SI_T、SI1、SI2、SI3、SI4、SI5、S1、S2、S3、SAIO、CRSI；植被指数包括：SAVI、GDVI。各光谱指数公式如表1所示。

本研究基于研究区影像，通过GEE云平台调用get Index函数构建光谱指数，选择红波段(Red)、绿波段(Green)、蓝波段(Blue)和近红外波段(NIR)，按照计算公式，进行各类光谱指数的计算和提取[33]，本文进行光谱指数提取操作程序代码均在GEE云平台在线代码编辑器中用JavaScript编写。

2.2 敏感光谱参数筛选

因自变量之间存在共线性问题，直接输入较多的变量建模会因样本数据之间的微小差异导致回归方程的回归系数出现误差，从而降低模型的稳定性，在一定程度上会影响建模精度[34]。为了消除输入变量过多和多重共线性在建模过程中带来的不利影响，本文基于R软件(版本4.1.3)，利用R语言中的Hmisc程序包中的cor函数进行不同数据源下SSC与13个光谱参数的Pearson相关性分析和显著性检验[44]。

表1 光谱指数计算公式

Tab. 1 The calculation formula of spectral indexes

光谱指数	计算公式	参考文献
盐分指数(SI_T)	(Red−NIR)×100	[35]
盐分指数1(SI1)	$\sqrt{Green\times Red}$	[36]
盐分指数2(SI2)	$\sqrt{Green^2+Red^2+NIR^2}$	[36]
盐分指数3(SI3)	$\sqrt{Green^2+Red^2}$	[37]
盐分指数4(SI4)	(Blue×Red)/Green	[38]
盐分指数5(SI5)	(Green+Red)/2	[39]
盐渍化指数1(S1)	Blue/Red	[38]
盐渍化指数2(S2)	(Blue−Red)/(Blue+Red)	[38]
盐渍化指数3(S3)	(Green×Red)/Blue	[38]
盐度比值指数(SAIO)	(Red−NIR)/(Green+NIR)	[40]
土壤调节植被指数(SAVI)	[(NIR−Red)×(1+L)]/(NIR+Red+L)；L=0.5	[41]
绿度差值植被指数(GDVI)	$(NIR^2-Red^2)/(NIR^2+Red^2)$	[42]
冠层响应盐度指数(CRSI)	$\sqrt{[(NIR\times Red)-(Green\times Blue)]/[(NIR\times Red)+(Green\times Blue)]}$	[43]

2.3 土壤盐渍化反演模型

2.3.1 随机森林模型构建 随机森林的核心思想是对训练集进行自助采样，组成多个训练集，每个训练集生成一棵决策树，所有决策树组成随机森林，从而对样本进行训练并预测的机器学习算法[45]。该方法将多个决策树组合起来以创建随机森林[46]，步骤如下：

步骤1：首先从训练样本集$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$中，随机产生$N$个训练集$S_1,S_2,\cdots,S_N$。

步骤2：对每个训练集生成对应的决策树$f_1,f_2,\cdots,f_N$；在对每个非叶节点上选择属性前，从全部M个属性中随机抽取m个($m<<M$)作为当前节点的分裂属性集，并从中选择出一个最佳分割属性作为节点进行分裂；

步骤3：将生成的多棵决策树构成随机森林，对于测试集样本X，利用每棵决策树进行测试，根据公式得到预测结果$f_1(x),f_2(x),\cdots,f_N(x)$，根据参考文献[47]，预测公式如下：

$$f(x)=\frac{1}{N}\sum_{n=1}^{N}f_N(x) \tag{2}$$

式中：$f(x)$表示N个$f_N(x)$的集合；$f(x)$表示随机森林模

型的预测结果;$f_N(x)$表示单个决策树的预测结果。

本文通过GEE云平台调用ee.Classifier.smileRandomForest函数实现模型回归预测,模型的效果通过决策树的数量n和结点使用的特征数m两个参数来调整。决策树数量过多会影响模型效率,过少则影响模型精度,因此在权衡决策树数量并兼顾精度和效率的同时,通过试验确定决策树数量为100,将结点使用的特征数设置为默认值,即输入特征总数的平方根。

2.3.2 模型评价指标 为量化土壤盐分反演模型的预测效果,利用GEE云平台计算并输出两个常用指标决定系数(R^2)和均方根误差(RMSE)来进行综合评价。R^2越接近于1,均方根误差RMSE越小,则模型拟合效果越好,预测效果越佳[48]。

3 结果与分析

3.1 土壤盐分统计特征

参考布雷迪的分级方法[49],将297份土样根据盐分含量将样本的盐渍化程度分为5个等级,样本的描述性统计分析如表2所示。由表2可知,不同盐渍化程度土壤样本数量大体均匀,土壤样本含盐量总体变异系数超过138%,变异性较强,表明样本离散程度较高,具有普适性。将土壤样本按含盐量由低到高排序,按照2:1的比例等间隔选取198个样本作为训练集建立反演模型,其余99个作为验证集,用于检验反演模型[16]。

3.2 土壤盐分与光谱指数相关性分析

利用R语言中的GGally和ggplot2程序包绘制不同数据源下SSC与13个光谱参数的相关矩阵图(图2),基于通过0.01显著性检验的敏感光谱参数建模。图2显示,在以Landsat8 OLI为数据源计算的光谱指数中,除了SI4、SAIO和GDVI,其余光谱指数与土壤含盐量均具有较强的相关性,SI_T、SI2、SI3和S3均通过0.01显著性检验,S1、S2和CRSI均通过0.001显著性检验,且S2与SSC的系数最大,表明光谱指数SI_T、SI2、SI3、S1、S2、S3和CRSI具有较强的相关性,可用于土壤盐分模型的构建。在以Sentinel-2A为数据源计算的光谱指数中,所有光谱指数与土壤含盐量均具有较强的相关性,SI1、SI3、SI5、SAIO、SAVI和GDVI均通过0.01显著性检验,SI_T、SI2、S1、S2、S3和CRSI均通过0.001显著性检验,表明SI_T、SI2、S1、S2、S3和CRSI与SSC有较强的相关性,故选择SI1、SI2、SI3、SI5、SAIO、SAVI、GDVI、SI_T、S1、S2、S3和CRSI用于以Sentinel-2A为数据源的模型构建。

3.3 模型构建

基于随机森林模型,以土壤含盐量为输出层,在以Landsat8 OLI为数据源建立的模型中,选取SI_T、SI2、SI3、S3、S1、S2和CRSI为输入层数据;在以Sentinel-2A为数据源建立的模型中,选取SI1、SI2、SI3、SI5、SAIO、SAVI、GDVI、SI_T、S1、S2、S3和CRSI为输入层数据,分别对土壤盐分含量进行模拟,结果如表3所示。由表3可知,以Sentinel-2A建立的模型的R^2值较以Landsat8 OLI建立的模型大,且RMSE小,表明以Sentinel-2A为数据源建立的土壤盐分预测模型较优。

利用99个验证数据对模型进行验证,结果如图3。图3表明以两种数据源建立模型,预测值与实测值均具有良好的相关性,其中以Sentinel-2A为数据源建立随机森林模型的R^2更大,且RMSE更小,表明以Sentinel-2A为数据源建立的土壤盐分随机森林预测模型精度优于以Landsat8 OLI为数据源建立的预测模型。

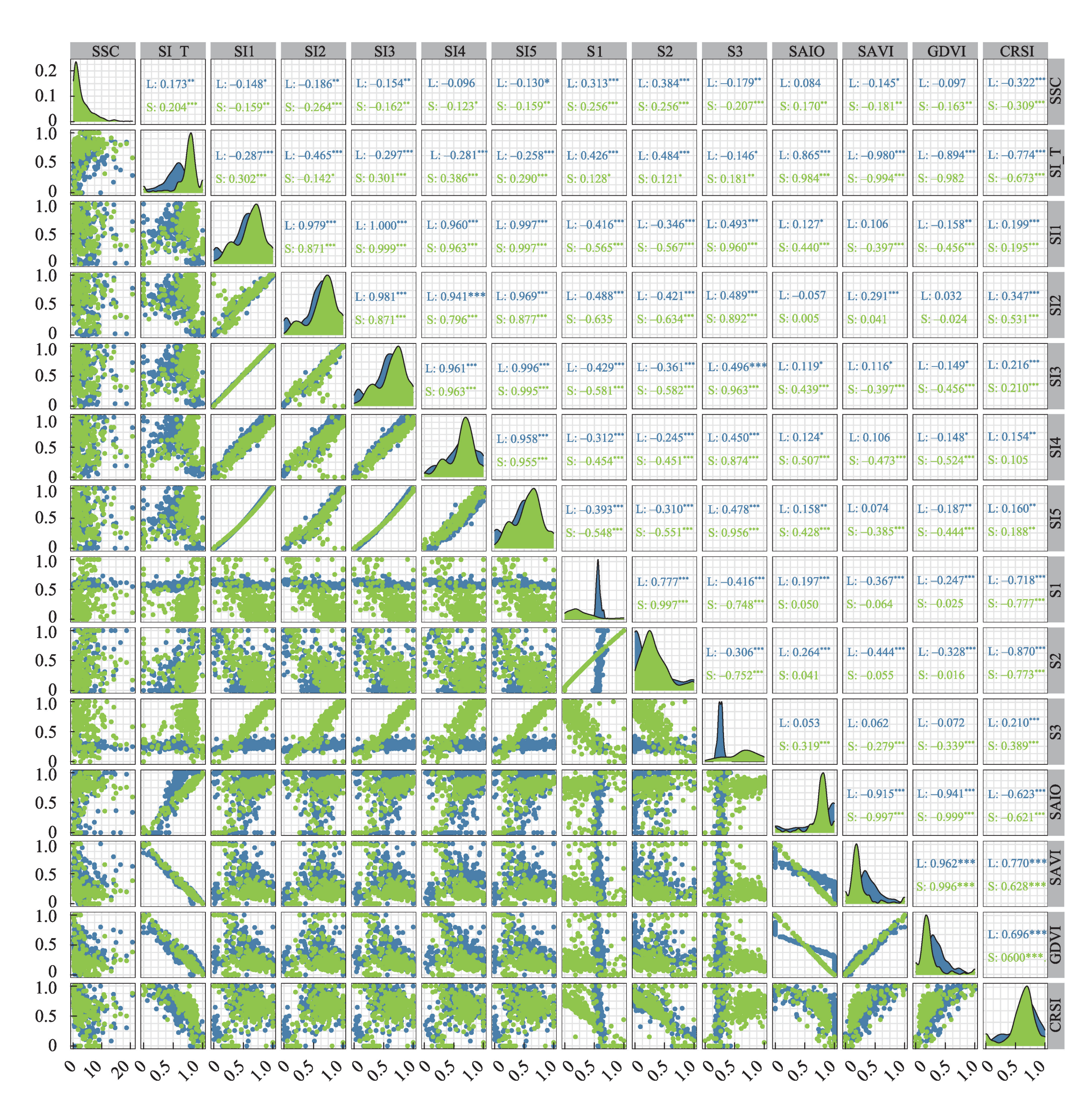
3.4 土壤盐渍化预测

利用以Sentinel-2A为数据源建立的土壤盐分随机森林预测模型,对银川平原土壤盐渍化进行预

表2 银川平原土壤样本统计

Tab. 2 Soil salinization degree and sample statistics in Yinchuan Plain

盐渍化等级	含盐量/($g\cdot kg^{-1}$)	样本数量	含盐量均值/($g\cdot kg^{-1}$)	含盐量最大值/($g\cdot kg^{-1}$)	含盐量最小值/($g\cdot kg^{-1}$)	变异系数/%
非盐渍化	<1	91	0.56	0.98	0.09	39.79
轻度盐渍化	1~2	64	1.46	1.96	1.00	20.09
中度盐渍化	2~4	55	2.82	3.97	2.00	19.42
重度盐渍化	4~6	37	4.90	5.91	4.06	11.97
盐土	>6	50	12.00	20.74	6.14	35.10
总样本		297	3.63	20.74	0.09	138.76



注：(1) *表示显著性检验 $P<0.05$，**表示显著性检验 $P<0.01$，***表示显著性检验 $P<0.001$；(2) 左下区为土壤含盐量(SSC)与13个光谱参数的两两关系散点图，右上区为对应的相关系数值，中间斜对角为对应横轴的核密度估计图；(3) 图中不同颜色代表不同数据源，蓝色代表以Landsat8 OLI为数据源计算得出的相关系数，绿色代表以Sentinel-2A为数据源计算得出的相关系数。

图2 光谱指数与土壤盐分的相关性

Fig. 2 Correlation between spectral indexes and soil salinity

表3 基于不同数据源的随机森林模型

Tab. 3 Random forest model based on different data sources

数据源	建模集(n=198)		验证集(n=99)	
	R^2	RMSE	R^2	RSME
Landsat8 OLI	0.615	3.169	0.734	1.698
Sentinel-2A	0.633	2.875	0.789	1.487

测，按照盐渍化程度对预测结果进行分级和统计，结果如图4和表4。由图4可知，银川平原土壤盐分总体上呈现南轻北重的分布格局，原因可能是平原地势南高北低，黄河流向自南向北，北部地下水埋深浅、盐分极易积累于土壤表层，盐土主要集中分布在平原北部的平罗县西侧、大武口区和惠农区，占盐土总面积的89.85%。南部作为黄河入水口携

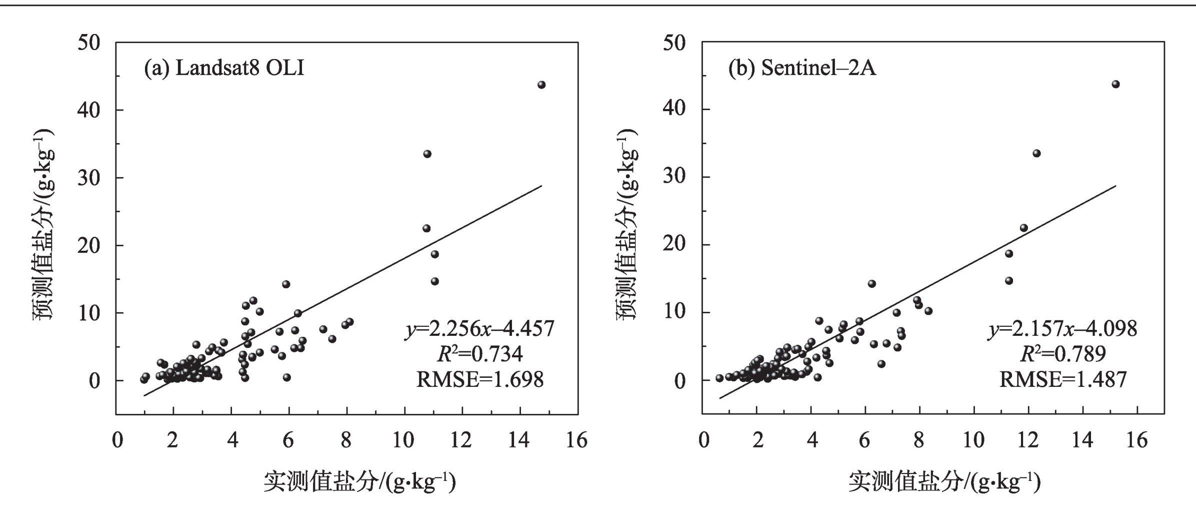


图3　土壤盐分与模型预测值的散点图

Fig. 3　The scatter plot of soil salinity and model predictions value

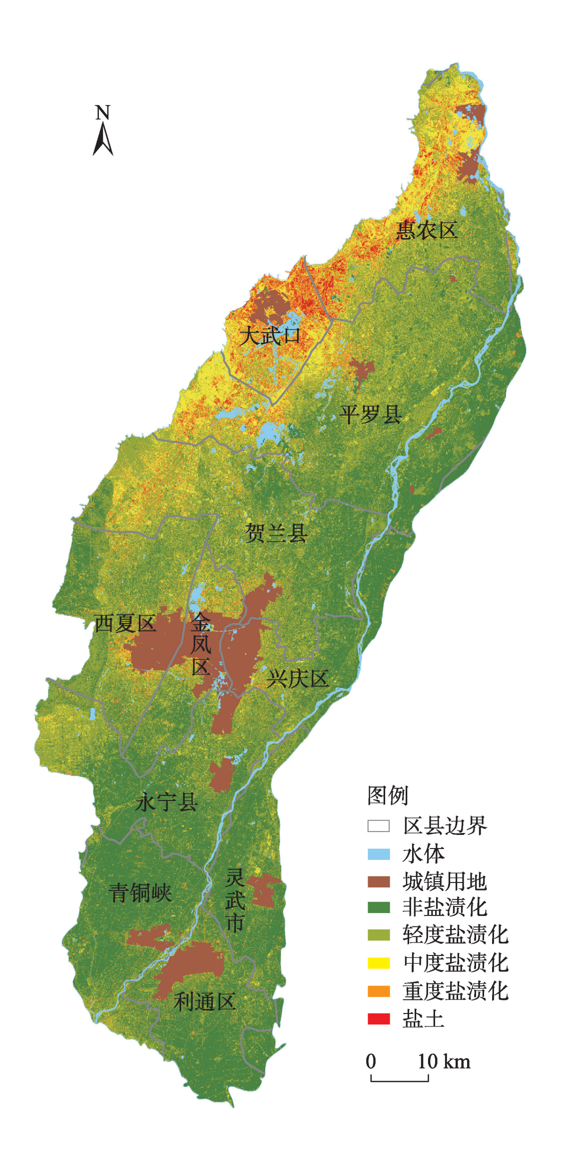


图4　土壤含盐量反演等级分布

Fig. 4　Grade distribution of soil salinity inversion

表4　银川平原不同等级盐分反演像元数及占比

Tab. 4　Number and proportion of inversion pixels of different grades of salinity in Yinchuan Plain

盐渍化等级	像元数	面积/km²	占比/%
非盐渍化	29115098	2916.701	41.620
轻度盐渍化	23348676	2332.925	33.290
中度盐渍化	10886719	1086.069	15.498
重度盐渍化	5121100	511.611	7.301
盐土	1609153	160.567	2.291

带盐基离子外流,土壤积盐较少,非盐渍化和轻度盐渍化土壤集中分布在平原南部的金凤区、兴庆区、永宁县、灵武市、青铜峡市和利通区,占银川平原总面积的71.59%。盐渍化预测结果与实际采样情况较为一致,表明随机森林模型可用于该区土壤盐分预测。

4　讨论

利用从卫星影像的各波段计算土壤光谱指数已成为监测土壤盐渍化空间分布的最常用方法[50],多数学者使用来自卫星数据可见光波段的土壤光谱指数,并选择与地面测量数据具有较高相关性的光谱指数作为建立模型的输入变量,均得到了较好的预测结果,表明建立光谱指数在土壤盐分监测应用方面具有一定的优势[51–52]。目前,常用的建模方案有两种:一是直接使用盐分指数估算土壤含盐量[53–54];二是利用植被指数间接估算土壤含盐量[55]。但为了

更加精确地监测土壤盐渍化,必须考虑盐生植物的光谱反射率,特别是在盐碱化严重的干旱和半干旱地区[56]。鉴于此,本文综合考虑了盐分指数和植被指数,通过Landsat8 OLI和Sentinel-2A的可见光波段和近红外波段信息建立多个光谱指数与盐分实测值进行相关性分析,使用随机森林算法建立回归模型监测盐渍土的空间分布,并在GEE平台上生成其结果。与常用的线性回归分析土壤盐分分布的方法相比,机器学习算法能够对土壤盐渍化空间分布提供更加可靠的预测效果,尤其是在实地测量数据有限的情况下[57]。而且与其他机器学习算法相比,随机森林算法能够更准确地预测盐渍化土地的土壤盐分分布情况[58]。

但当前研究还具有许多不确定性,主要有两个来源:一是土壤盐分特征与遥感数据之间关系的不确定性。由于土壤盐渍化的机制相当复杂,通过卫星图像获取的土壤盐分信息可能会受到土壤类型、土壤表面粗糙度和其他土壤特性的影响[59]。其根本原因是卫星传感器在信息获取方面存在先天缺陷,不能表达出土壤的定性指标。此外,虽然遥感影像已经过辐射校正和大气校正,但捕获的光谱信息仍可能受到地形因素的影响。二是建立随机森林模型时的不确定性,对于机器学习模型来说,数据集的质量和数量对模型的预测结果有很大的影响,本研究中样本量(n=297)相对较小,这可能会在建立随机森林模型时产生不确定性。因此,本文所建立的估算模型能否运用到其他地区还有待进一步的研究和验证。

5 结 论

在本研究中,通过比较Landsat8 OLI和Sentinel-2A图像的光谱特征,并利用两幅不同遥感影像建立随机森林模型研究对比了土壤含盐量的预测性能,主要研究结果如下:

(1) 利用GEE云计算平台提供的来自不同传感器的大量分析数据和算法,有助于准确估计研究区的土壤盐分分布情况,促进大规模区域土壤盐分监测工作的进行。

(2) 基于Sentinel-2A数据建立的随机森林模型表现出了较好的预测精度和模型性能(R^2=0.789,RMSE=1.487),土壤含盐量预测结果优于以Landsat8 OLI为数据源建立的随机森林模型。

(3) 利用基于Sentinel-2A数据的随机森林模型对银川平原空间范围内土壤盐分进行预测,预测结果与实地采样情况具有较好的一致性。表明以Sentinel-2A为数据源建立的随机森林模型可用于土壤盐渍化预测。

参考文献(References):

[1] 李星, 李凯锋. 土壤盐渍化开发利用研究进展[J]. 农业科技通讯, 2018, 23(6): 244-246. [Li Xing, Li Kaifeng. Research progress of soil salinization development and utilization[J]. Agricultural Science and Technology Bulletin, 2018, 23(6): 244-246.]

[2] Nachshon U. Cropland soil salinization and associated hydrology: Trends, processes and examples[J]. Water, 2018, 10(8): 1030.

[3] Zaman M, Shahid S A, Heng L. Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques [M]. Springer Nature, 2018.

[4] 杨真, 王宝山. 中国盐渍土资源现状及改良利用对策[J]. 山东农业科学, 2015, 47(4): 125-130. [Yang Zhen, Wang Baoshan. Present situation of saline soil resources in China and its improvement and utilization strategy[J]. Shandong Agricultural Sciences, 2015, 47(4): 125-130.]

[5] 曹雷, 丁建丽, 玉米提·哈力克, 等. 基于国产高分一号卫星数据的区域土壤盐渍化信息提取与建模[J]. 土壤学报, 2016, 53(6): 1399-1409. [Cao Lei, Ding Jianli, Coranti Hallick, et al. Extraction and modeling of regional soil salinization based on domestic Gaofen-1 satellite data[J]. Acta Pedologica Sinica, 2016, 53(6): 1399-1409.]

[6] Meng L, Zhou S, Zhang H, et al. Estimating soil salinity in different landscapes of the Yellow River Delta through Landsat OLI/TIRS and ETM+ Data[J]. Journal of Coastal Conservation, 2016, 20(4): 271-279.

[7] Farifteh J, Farshad A, George R J. Assessing salt-affected soils using remote sensing, solute modelling, and geophysics[J]. Geoderma, 2006, 130(3-4): 191-206.

[8] Hu J, Peng J, Zhou Y, et al. Quantitative estimation of soil salinity using UAV-borne hyperspectral and satellite multispectral images [J]. Remote Sensing, 2019, 11(7): 736.

[9] Metternicht G I, Zinck J A. Remote sensing of soil salinity: Potentials and constraints[J]. Remote sensing of Environment, 2003, 85(1): 1-20.

[10] Fan X, Weng Y, Tao J. Towards decadal soil salinity mapping using Landsat time series data[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 52: 32-41.

[11] Ivushkin K, Bartholomeus H, Bregt A K, et al. Global mapping of soil salinity change[J]. Remote Sensing of Environment, 2019, 231: 111-160.

[12] 黄晓宇, 王雪梅, 卡吾恰提·白山. 基于Landsat8 OLI影像干旱区绿洲土壤含盐量反演[J]. 自然资源遥感, 2023, 35(1): 189–197. [Huang Xiaoyu, Wang Xuemei, Kawuciati Baishan. Inversion of soil salt content in arid oasis based on Landsat8 OLI image[J]. Remote Sensing of Natural Resources, 2023, 35(1): 189–197.]

[13] 杨宁, 崔文轩, 张智韬, 等. 无人机多光谱遥感反演不同深度土壤盐分[J]. 农业工程学报, 2020, 36(22): 13–21. [Yang Ning, Cui Wenxuan, Zhang Zhitao, et al. Inversion of soil salinity at different depths by UAV multi-spectral remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(22): 13–21.]

[14] 陈俊英, 姚志华, 张智韬, 等. 大田葵花土壤含盐量无人机遥感反演研究[J]. 农业机械学报, 2020, 51(7): 178–191. [Chen Junying, Yao Zhihua, Zhang Zhitao, et al. Uav remote sensing inversion of soil salt content in sunflower of Dantian[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 178–191.]

[15] 厉彦玲, 赵庚星, 常春艳, 等. OLI与HSI影像融合的土壤盐分反演模型[J]. 农业工程学报, 2017, 33(21): 533–544. [Li Yanling, Zhao Gengxing, Chang Chunyan, et al. Soil salinity inversion model based on OLI and HSI image fusion[J]. Transactions of the CSAE, 2017, 33(21): 533–544.]

[16] 张智韬, 魏广飞, 姚志华, 等. 基于无人机多光谱遥感的土壤含盐量反演模型研究[J]. 农业机械学报, 2019, 50(12): 151–160. [Zhang Zhitao, Wei Guangfei, Yao Zhihua, et al. Inversion model of soil salt content based on UAV multi-spectral remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 151–160.]

[17] Fathizad H, Ardakani M A H, Sodaiezadeh H, et al. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran[J]. Geoderma, 2020, 365: 114–233.

[18] Liu X, Hu G, Chen Y, et al. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine platform[J]. Remote Sensing of Environment, 2018, 209: 227–239.

[19] Simonetti D, Simonetti E, Szantoi Z, et al. First results from the phenology-based synthesis classifier using Landsat 8 imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(7): 1496–1500.

[20] 徐晗泽宇, 刘冲, 王军邦, 等. Google Earth Engine平台支持下的赣南柑橘果园遥感提取研究[J]. 地球信息科学学报, 2018, 20(3): 396–404. [Xu Hanzeyu, Liu Chong, Wang Junbang, et al. Remote sensing extraction of citrus orchards in Southern Jiangxi Province supported by Google Earth Engine[J]. Journal of Geoinformation Science, 2018, 20(3): 396–404.]

[21] Dong J, Xiao X, Menarguez M A, et al. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine[J]. Remote Sensing of Environment, 2016, 185: 142–154.

[22] Lobell D B, Thau D, Seifert C, et al. A scalable satellite-based crop yield mapper[J]. Remote Sensing of Environment, 2015, 164: 324–333.

[23] Tamiminia H, Salehi B, Mahdianpari M, et al. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 164: 152–170.

[24] Hansen M C, Potapov P V, Moore R, et al. High-resolution global maps of 21st-century forest cover change[J]. Science, 2013, 342 (6160): 850–853.

[25] Pekel J F, Cottam A, Gorelick N, et al. High-resolution mapping of global surface water and its long-term changes[J]. Nature, 2016, 540(7633): 418–422.

[26] Trianni G, Lisini G, Angiuli E, et al. Scaling up to national/regional urban extent mapping using Landsat data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(7): 3710–3719.

[27] 张源沛, 胡克林, 李保国, 等. 银川平原土壤盐分及盐渍土的空间分布格局[J]. 农业工程学报, 2009, 25(7): 19–24. [Zhang Yuanpei, Hu Keling, Li Baoguo, et al. Spatial distribution pattern of soil salinity and saline soil in Yinchuan Plain[J]. Transactions of the Chinese Society of Agricultural Engineering, 2009, 25(7): 19–24.]

[28] 何文寿, 刘阳春, 何进宇. 宁夏不同类型盐渍化土壤水溶盐含量与其电导率的关系[J]. 干旱地区农业研究, 2010, 28(1): 111–116. [He Wenshou, Liu Yangchun, He Jinyu. Relationship between water-soluble salt content and electrical conductivity of different salinized soils in Ningxia[J]. Agricultural Research in the Arid Areas, 2010, 28(1): 111–116.]

[29] 肖文娟. 基于GEE的近10年来云南省耕地变化遥感监测研究[D]. 昆明: 云南师范大学, 2020. [Xiao Wenjuan. Remote Sensing Monitoring of Cultivated Land Change in Yunnan Province in Recent 10 Years Based on GEE[D]. Kunming: Yunnan Normal University, 2020.]

[30] 丁建丽, 伍漫春, 刘海霞, 等. 基于综合高光谱指数的区域土壤盐渍化监测研究[J]. 光谱学与光谱分析, 2012, 32(7): 1918–1922. [Ding Jianli, Wu Manchun, Liu Haixia, et al. Monitoring of regional soil salinization based on comprehensive hyperspectral index[J]. Spectroscopy and Spectral Analysis, 2012, 32(7): 1918–1922.]

[31] 贾萍萍, 尚天浩, 张俊华, 等. 利用多源光谱信息反演宁夏银北地区干湿季土壤含盐量[J]. 农业工程学报, 2020, 36(17): 125–134. [Jia Pingping, Shang Tianhao, Zhang Junhua, et al. Inversion of soil salt content in wet and dry seasons in Yinbei Region of Ningxia using multi-source spectral information[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(17): 125–134.]

[32] 杨小虎, 罗艳琴, 杨海昌, 等. 玛纳斯河流域绿洲农田土壤盐分

反演及空间分布特征[J]. 干旱区资源与环境, 2021, 35(2): 156–161. [Yang Xiaohu, Luo Yanqin, Yang Haichang, et al. Inversion and spatial distribution of soil salinity in oasis farmland in Manas River Basin[J]. Journal of Arid Land Resources and Environment, 2021, 35(2): 156–161.]

[33] 刘旭辉, 白云岗, 柴仲平, 等. 基于多光谱遥感的典型绿洲棉田春季土壤盐分反演及验证[J]. 干旱区地理, 2022, 45(4): 1165–1175. [Liu Xuhui, Bai Yungang, Chai Zhongping, et al. Inversion and verification of soil salinity in spring in typical oasis cotton fields based on multispectral remote sensing[J]. Arid Land Geography, 2022, 45(4): 1165–1175.]

[34] 陈俊英, 王新涛, 张智韬, 等. 基于无人机–卫星遥感升尺度的土壤盐渍化监测方法[J]. 农业机械学报, 2019, 50(12): 161–169. [Chen Junying, Wang Xintao, Zhang Zhitao,et al. Soil salinization monitoring method based on UAV-Satellite remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(12): 161–169.]

[35] Allbed A, Kumar L, Aldakheel Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region[J]. Geoderma, 2014, 230: 1–8.

[36] 樊彦国, 张维康, 刘敬一. 基于植被指数–盐分指数特征空间的黄河三角洲盐渍化遥感监测研究[J]. 山东农业科学, 2016, 48(5): 137–141. [Fan Yanguo, Zhang Weikang, Liu Jingyi. Remote sensing monitoring of salinization in Yellow River Delta based on vegetation index and salinity index feature space[J]. Shandong Agricultural Sciences, 2016, 48(5): 137–141.]

[37] 姚远, 丁建丽, 张芳, 等. 基于高光谱指数和电磁感应技术的区域土壤盐渍化监测模型[J]. 光谱学与光谱分析, 2013, 33(6): 1658–1664. [Yao Yuan, Ding Jianli, Zhang Fang, et al. Regional soil salinization monitoring model based on hyperspectral index and electromagnetic induction technology[J]. Spectroscopy and Spectral Analysis, 2013, 33(6): 1658–1664.]

[38] Khan S, Abbas A. Using remote sensing techniques for appraisal of irrigated soil salinity[J]. Int. Congr. Model. Simul.(MODSIM), Model. Simul. Soc. Aust. New Zealand, Bright, 2007 (January): 2632–2638.

[39] Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data[J]. Geoderma, 2006, 134(1–2): 217–230.

[40] 王飞, 杨胜天, 丁建丽, 等. 环境敏感变量优选及机器学习算法预测绿洲土壤盐分[J]. 农业工程学报, 2018, 34(22): 102–110. [Wang Fei, Yang Shengtian, Ding Jianli, et al. Optimization of environmental sensitive variables and prediction of oasis soil salinity by machine learning algorithm[J]. Transactions of the CSAE, 2018, 34(22): 102–110.]

[41] Huete A R. A soil-adjusted vegetation index (SAVI)[J]. Remote Sensing of Environment, 1988, 25(3): 295–309.

[42] Wu W. The generalized difference vegetation index (GDVI) for dryland characterization[J]. Remote Sensing, 2014, 6(2): 1211–1233.

[43] Scudiero E, Skaggs T H, Corwin D L. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA[J]. Geoderma Regional, 2014, 2: 82–90.

[44] Parolo G, Rossi G, Ferrarini A. Toward improved species niche modelling: Arnica montana in the Alps as a case study[J]. Journal of Applied Ecology, 2008, 45(5): 1410–1418.

[45] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.

[46] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014. [Cao Zhengfeng. Research on Optimization of Random Forest Algorithm[D]. Beijing: Capital University of Economics and Business, 2014.]

[47] Gieseke F, Igel C. Training big random forests with little resources [C]//Proceedings of the 24th Acmsigkdd International Conference on Knowledge Discovery & Data Mining, 2018: 1445–1454.

[48] 冯娟, 丁建丽, 杨爱霞, 等. 干旱区土壤盐渍化信息遥感建模[J]. 干旱地区农业研究, 2018, 36(1): 266–273. [Feng Juan, Ding Jianli, Yang Aixia, et al. Remote sensing modeling of soil salinization in arid areas[J]. Agricultural Research in the Arid Areas, 2018, 36(1): 266–273.]

[49] 武丹, 贾科利, 张晓东, 等. 基于异质SVM神经网络的土壤盐渍化灾害预测模型[J]. 水文地质工程地质, 2018, 45(5): 143–149, 157. [Wu Dan, Jia Keli, Zhang Xiaodong, et al. Soil salinization disaster prediction model basedon heterogeneous SVM neural network[J]. Hydrogeology and Engineering Geology, 2018, 45(5): 143–149, 157.]

[50] Morgan R S, El-Hady M A, Rahim I S. Soil salinity mapping utilizing sentinel–2 and neural networks[J]. Indian Journal of Agricultural Research, 2018, 52(5): 524–529.

[51] Allbed A, Kumar L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review [J]. Advances in Remote Sensing, 2013, 2: 373–385.

[52] Gorji T, Yildirim A, Hamzehpour N, et al. Soil salinity analysis of Urmia Lake Basin using Landsat–8 OLI and Sentinel–2A based spectral indices and electrical conductivity measurements[J]. Ecological Indicators, 2020, 112: 106–173.

[53] Han L, Liu D, Cheng G, et al. Spatial distribution and genesis of salt on the saline playa at Qehan Lake, Inner Mongolia, China[J]. Catena, 2019, 177: 22–30.

[54] Wang J, Ding J, Yu D, et al. Capability of Sentinel–2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China[J]. Geoderma, 2019, 353: 172–187.

[55] Zhang T T, Qi J G, Gao Y, et al. Detecting soil salinity with MODIS time series VI data[J]. Ecological Indicators, 2015, 52: 480–489.

[56] Scudiero E, Skaggs T H, Corwin D L. Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance[J]. Remote

Sensing of Environment, 2015, 169: 335-343.

[57] 王明宽, 莫宏伟, 陈红艳. 基于多光谱影像反演土壤盐分的建模方法研究[J]. 土壤通报, 2016, 47(5): 1036-1041. [Wang Mingkuan, Mo Hongwei, Chen Hongyan. Modeling method of soil salinity inversion based on multi-spectral image[J]. Chinese Journal of Soil Science, 2016, 47(5): 1036-1041.]

[58] 杨练兵, 陈春波, 郑宏伟, 等. 基于优化随机森林回归模型的土壤盐渍化反演[J]. 地球信息科学学报, 2021, 23(9): 1662-1674. [Yang Lianbing, Chen Chunbo, Zheng Hongwei, et al. Soil salinization inversion based on optimized stochastic forest regression model[J]. Journal of Geo- Information Science, 2021, 23(9): 1662-1674.]

[59] Ding J, Yu D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments [J]. Geoderma, 2014, 235: 316-322.

Prediction of soil salt content based on the random forest algorithm

LI Xiaoyu, JIA Keli, WEI Huimin, CHEN Ruihua, WANG Yijing

(College of Geographical Sciences and Planning, Ningxia University, Yinchuan 750021, Ningxia, China)

Abstract: Soil salinization caused by natural and anthropogenic factors is an environmental hazard that is especially important in arid and semi-arid regions of the world. The accumulation of salts in soil is a major threat to crop production and global agriculture. Therefore, the rapid and precise detection of salt- affected lands is highly critical for sustaining soil productivity. This paper aims to analyze the performance of the random forest algorithm in mapping soil salinity in the Yinchuan Plain using Landsat-8 OLI, Sentinel-2A satellite images, and ground-based soil salt content (SSC) measurements with the aid of the Google Earth Engine (GEE) platform. We estimated SSC by establishing the relationship between spectral index characteristics and ground-measured soil salt content. The results show that GEE can provide reliable data support for soil salinity prediction. The random forest model established with Sentinel-2A as the data source performed better (R^2 = 0.789, RMSE = 1.487) than and can therefore be used for the estimation of soil salinity using high- resolution remote sensing, which can provide theoretical support for large-scale soil salinity monitoring.

Keywords: soil salt content; Google Earth Engine; random forest; predict; Yinchuan Plain